

SITE SEARCH



Method and apparatus for branch prediction using second level branch prediction table

Application Number	99813061	Application Date	1999.08.26
Publication Number	1328664	Publication Date	2001.12.26
Priority Information	US09/1498851998/9/8		
International Classification	G06F9/30		
Applicant(s) Name	Intel Corp.		
Address			
Inventor(s) Name	T.-Y. Yeh;H.P. Sharangpani		
Patent Agency Code	72001	Patent Agent	wu zengyong

Abstract

A branch predictor. A first branch prediction table (100) is coupled to an instruction pointer generator (11) to store tagged branch prediction entries and to provide branch predictions at high speed. A second branch prediction table (102) is also coupled to the instruction pointer generator (11) to store untagged branch prediction entries and to provide branch predictions for a much larger working set of branches, albeit at a slower speed.



[12] 发明专利申请公开说明书

[21] 申请号 99813061.3

[43] 公开日 2001 年 12 月 26 日

[11] 公开号 CN 1328664A

[22] 申请日 1999.8.26 [21] 申请号 99813061.3

[30] 优先权

[32] 1998.9.8 [33] US [31] 09/149,885

[86] 国际申请 PCT/US99/19892 1999.8.26

[87] 国际公布 W000/14628 英 2000.3.16

[85] 进入国家阶段日期 2001.5.8

[71] 申请人 英特尔公司

地址 美国加利福尼亚州

[72] 发明人 T·Y·叶 H·P·沙朗帕尼

[74] 专利代理机构 中国专利代理(香港)有限公司

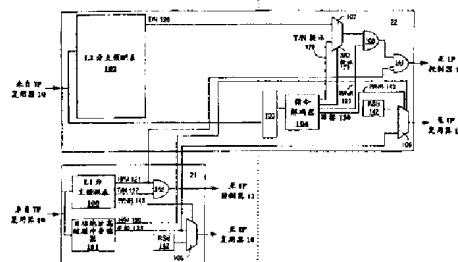
代理人 吴增勇 王忠忠

权利要求书 4 页 说明书 12 页 附图页数 3 页

[54] 发明名称 利用第二级分支预测表的分支预测的方法和设备

[57] 摘要

一种分支预测器。第一分支预测表(100)连接到指令指针发生器(11)以存储标记的分支预测入口并高速地提供分支预测。第二分支预测表(102)也连到指令指针发生器(11)以存储未标记的分支预测入口并提供大得多的分支工作集的分支预测,尽管是以较慢速度进行提供。



I S S N 1 0 0 8 - 4 2 7 4

权 利 要 求 书

1. 一种分支预测器包括:

5 连至指令指针(IP)发生器的第一分支预测表(BPT), 用于存储多个分支预测入口; 和

连至所述 IP 发生器的第二 BPT, 用于存储更多个分支预测入口。

2. 权利要求 1 的分支预测器, 其特征在于所述第一 BPT 存储已标记的分支预测入口而所述第二 BPT 存储未标记的分支预测入口。

3. 权利要求 1 的分支预测器, 其特征在于还包括连至所述第一
10 BPT 的第一电路, 如果所述第一 BPT 表示产生分支, 则所述第一电路将 IP 地址预测为目标地址高速缓冲存储器中存储的目标地址。

4. 权利要求 1 的分支预测器, 其特征在于还包括连至所述第二 BPT 的第二电路, 如果所述第二 BPT 表示产生分支, 则所述第二电路将 IP 地址预测为目标地址高速缓冲存储器中存储的目标地址。

15 5. 权利要求 1 的分支预测器, 其特征在于还包括连至所述第二 BPT 的第二电路, 如果所述第二 BPT 表示产生分支, 则所述第二电路将 IP 地址预测为从指令解码及计算的目标地址。

6. 权利要求 1 的分支预测器, 其特征在于还包括连至所述第二 BPT 的第二电路, 如果所述第二 BPT 表示产生分支, 则所述第二电路
20 将 IP 地址预测为返回堆栈缓冲器中存储的目标地址。

7. 权利要求 1 的分支预测器, 其特征在于把所述第二 BPT 的大小做成存储两倍于或更多倍于所述第一 BPT 存储的分支预测入口数。

8. 权利要求 1 的分支预测器, 其特征在于所述第一 BPT 执行第
25 一种预测算法而所述第二 BPT 执行不同于所述第一种预测算法的第二种预测算法。

9. 一种处理器包括:

指令指针(IP)发生器;

连至所述 IP 发生器的第一级分支预测表(BPT), 所述第一级 BPT

其中存储有与第一多个地址有关的第一多个分支预测入口；和

连至所述 IP 发生器的第二级 BPT，所述第二级 BPT 其中存储有与第二多个地址有关的第二多个分支预测入口。

10. 权利要求 9 的处理器，其特征在于还包括连至所述 IP 发生器的目标地址高速缓冲存储器并在其中存储有与所述第一多个地址有关的多个目标地址。

11. 权利要求 10 的处理器，其特征在于还包括第一电路，如果在所述第一级 BPT 中有表示产生分支的命中，则所述第一电路从所述目标地址高速缓冲存储器中选择目标地址作为其输出。

12. 权利要求 9 的处理器，其特征在于还包括连至所述 IP 发生器的存储器单元中的指令组并在其中存储有多个分支提示。

13. 权利要求 12 的处理器，其特征在于还包括第二电路，如果在所述第二级 BPT 中的命中表示产生分支并且在所述存储器单元中的有关提示是动态的，则所述第二电路从所述存储器单元中选择目标地址作为其输出。

14. 权利要求 9 的处理器，其特征在于：在所述第二多个地址中有像在所述第一多个地址中两到四倍之间那么多的地址，并且所述第二多个地址包含了所述第一多个地址。

15. 一种在处理器中执行分支预测的方法，包括以下步骤：
20 搜索第一级分支预测表(BPT)以查找与指令指针(IP)地址有关的第一级分支预测入口；

搜索第二级 BPT 以查找与所述 IP 地址有关的第二级分支预测入口；和

如果在所述第一级 BPT 中发现所述第一分支预测入口，则基于所述第一级分支预测入口进行分支预测，并且，如果在所述第一级 BPT 中所述第一分支预测入口未命中，则基于所述第二分支预测入口进行分支预测。

16. 权利要求 15 的分支预测器，其特征在于还包括如下步骤：在

所述第一级BPT中执行第一种分支算法而在所述第二级BPT中执行与
所述第一种预测算法不同的第二种预测算法。

17. 权利要求 15 的方法，其特征在于如下步骤：如果与所述 IP
地址有关的指令被解码并被确为分支指令，则基于所述第二分支预测
5 入口进行分支预测。

18. 权利要求 15 的方法，其特征在于：搜索所述第一级 BPT 和
搜索所述第二级 BPT 的所述步骤同时发生。

19. 权利要求 15 的方法，其特征在于：搜索所述第一级 BPT 的
所述步骤包括将所述 IP 地址的地址标记与存储在所述第一级 BPT 中
10 的地址标记进行比较的步骤，搜索所述第二级 BPT 的步骤包括从直接
映射表中选择入口的步骤。

20. 一种在给定了初始 IP 地址的处理器中预测后继 IP 地址的方
法，所述方法包括以下步骤：

如果在第一级分支预测表(BPT)中与所述初始 IP 地址有关的分支
15 预测入口表示产生分支，则将所述后继 IP 地址预测为来自目标地址高
速缓冲存储器的目标地址；并且

如果在所述第一 BPT 中没有与所述初始 IP 地址有关的分支预测
入口并且在第二 BPT 中与所述初始 IP 地址有关的分支预测入口表示
产生分支，则将所述后继 IP 地址预测为从与所述初始 IP 地址有关的
20 指令解码的目标地址。

21. 权利要求 20 的方法，其特征在于还包括下述步骤：如果在所
述第一 BPT 中与所述初始 IP 地址有关的分支预测入口表示不产生分
支，则将所述后继 IP 地址预测为递增一预定量的所述初始 IP 地址。

22. 权利要求 20 的方法，其特征在于还包括下述步骤：如果在所
25 述第一 BPT 中没有与所述初始 IP 地址有关的分支预测入口并且在第
二 BPT 中与所述初始 IP 地址有关的分支预测入口表示不产生分支，
则将所述后继 IP 地址预测为递增一预定量的所述初始 IP 地址。

23. 权利要求 20 的方法，其特征在于还包括下述步骤：如果在所

述第一 BPT 中没有与所述初始 IP 地址有关的分支预测入口、第二 BPT 中与所述初始 IP 地址有关的分支预测入口表示不产生分支，且与所述初始 IP 地址有关的指令提示被解码为动态的，则将所述后继 IP 地址预测为递增一预定量的所述初始 IP 地址。

- 5 24. 权利要求 20 的方法，其特征在于将所述后继 IP 地址预测为来自目标地址高速缓冲存储器中的目标地址的所述步骤包括搜索所述第一 BPT 以查找同与所述初始 IP 地址有关的地址标记相匹配的地址标记的步骤，将所述后继 IP 地址预测为从指令解码的目标地址的所述步骤包括在所述第二 BPT 中选择与所述初始 IP 地址有关的直接映射
- 10 入口的步骤。

说明书

利用第二级分支预测表的分支预测的方法和设备

5 发明领域

本发明涉及计算机系统，更具体地说，涉及使用第一级和第二级分支预测表进行分支预测的处理器。

发明背景

10 先进的处理器采用流水(管道)线技术高速地执行指令。基于这样的处理器，整个机器组织成包括几个硬件级联级的多流水线。指令处理分为一系列的操作，且由相应的流水线级(管道级)的硬件来执行各种操作。来自几条指令的独立操作可由不同的流水线级同时进行，以增加处理器的指令通过量。当一个流水线处理器在每个流水线级
15 包括多个执行资源时，该处理器的指令通过量就能超过每个时钟周期一条指令。为了充分利用这个指令执行能力，必须由正确的执行路径向该处理器的执行资源提供足够的指令。

在典型的计算机系统中，指令指针(IP)指示处理器从程序代码的一条指令到下一条指令。一条指令可以正常的程序代码顺序指示这个 IP 到下一条指令，或者指示该 IP 跳过部分程序代码并继续执行非
20 正常顺序的指令。这个使处理器或者按顺序连续执行下一条指令或者“转移”到一个不同的非顺序的指令的指令称为分支指令。

例如，字处理器在进行拼写检查时，执行软件指令来确认正确地拼写了每个字。只要字拼写正确，指令便顺序地执行。然而一旦
25 发现了拼写有误的字，分支指令便指示 IP 转向一个子程序，后者提醒用户拼字错误。然后，处理器执行该子程序。

分支指令向保持流水线充满来自正确执行路径的指令这一要求提出了挑战。如果执行分支指令并满足分支条件，处理器的控制流



就跳向新的代码序列，来自该新的代码序列的指令转向流水线。分支执行一般出现在流水线的后端，而在流水线的前端取指令。如果获取指令依赖分支执行来决定正确的执行路径，则在决定分支条件前，处理器的流水线可能充满来自错误执行路径的指令。因此，必须从流水线中清洗（flush）这些指令，使得获取来自正确执行路径的指令时这些受到影响的流水线级的资源保持空闲。空闲的流水线级称为流水线泡(bubble)，因为直到它们被来自正确执行路径的指令填充时才提供有用的输出。

先进的处理器在它们流水线的前端加入分支预测模块以减少流水线泡的数量。当分支指令进入到流水线的前端时，分支预测模块预测在流水线的后端执行过程中是否产生该分支指令。如果预测产生分支(非顺序执行指令)，则分支预测模块就向指令提取模块提供分支目标地址，通过设置与含有分支程序代码的第一指令的地址相同的 IP 地址来重新指示 IP。含有分支代码的所述第一指令的地址称为“目标地址”。同样位于流水线前端的提取模块从目标地址开始取指令。另一方面，如果分支预测器预测不产生分支(顺序指令执行)，分支预测器就递增 IP 地址，以便 IP 以正常程序代码顺序指向下一条指令。当分支执行出现在流水线的后端时，处理器能够确认前端所做的预测是否正确。如果不正确，流水线就被清洗。分支预测准确性越高，流水线泡和清洗数就越少。

传统的分支预测模块采用分支目标缓冲器(BTB)来存储预测入口，后者包含诸如是否产生分支的信息以及产生分支时可能的目标地址。这些分支预测入口与含有分支指令的 IP 地址相联系。对每一个在分支预测表中被跟踪的 IP 地址，其有关的分支预测入口包含 IP 地址以及有助于预测将来是否产生分支的历史信息。然而，即使是在 BTB 中寻找指令、确定是否产生分支以及在已做预测的基础上向取指模块提供目标地址这样的过程也会在使处理器重转向(resteer)目标地址时引起延迟。该延迟允许来自错误执行路径的指令进入并沿

流水线传播。由于这些指令没有按预测的执行路经加到正向进程，在清洗它们时它们就会在流水线上形成“泡”。更准确且更完备的分支预测算法(用更大的分支表)需要更长的时间来完成并会在重转向过程中产生更大的延迟。重转向流水线需要的时钟周期数越大，在流水线中产生的泡数就越多。因此，在进入分支预测结构的速度与这些结构中内容的大小和准确性之间存在折中。

由于速度和费用的原因，现代处理器常限制所采用的 BTB 的容量。特别是在大工作负荷时，这降低了分支检测和预测的准确性。假如 BTB 的容量较小，则某个新分支预测入口有时必须重写某个旧分支预测入口。如果与重写的分支预测入口相联系的分支指令之后由处理器重执行，则没有历史信息来帮助分支预测器预测是否应该产生分支。结果，分支预测准确性下降，降低了处理器的性能。随着软件应用程序规模增大，这些应用程序中分支指令数目增加，分支预测表的有限容量就成为一个重要的问题。因此，有必要提供一种解决方案：为最频繁出现的分支子集(具有高局部性的那些)产生小的等待时间并为整个工作集提供有意义的预测。

发明概述

描述了一种分支预测器。第一级分支预测表与 IP 发生器连接以存储分支预测入口。第二级分支预测表也与 IP 发生器连接以存储更多分支预测入口。

按照本发明实施例，尽管降低了速度，然而可以发现两级分支预测结构充分地将高速(小的等待时间)分支预测和分支的最高局部性的重转向的益处与全工作集的全局高准确分支检测和预测结合在一起。这在没有显著增加小片尺寸情况下也可完成。

通过附图和以下的详细描述，本发明的其它特征和优点将一清二楚。

附图简介

通过附图中的示例而非限制来说明本发明，图中用相同标号表示类似的单元。附图中：

图 1 是一个表示本发明的方法的流程图；

5 图 2 表示按照本发明实施例的分支预测流水线；和

图 3 表示图 2 的分支预测器。

详细描述

10 描述了一种用于带有第一级和第二级分支预测表的处理器的分支预测器。IP 发生器产生初始指令指针(IP)地址。为与初始 IP 地址相关联的分支预测入口搜索第一级(L1) BPT(分支预测表)和第二级(L2)BPT。对本发明的一个实施例，L1 BPT 是相联的(即完全相联或多路集相联)而 L2 BPT 则是直接映射。

15 因为 L1 BPT 是关联的，所以分支预测入口加上标记并将这些标记与预测信息一起存于 L1 BPT 中。未对直接映射的 L2 BPT 中的分支预测入口作标记，故 L2 BPT 中只存有分支预测信息。因为 L1 BPT 中的分支预测入口大于 L2 BPT 中的分支预测入口，在等量空间中，L2 BPT 可比 L1 BPT 有更多的分支预测入口。因此尽管 L1 BPT 可能比 L2 BPT 更快更准确，然而 L2 BPT 作为 L1 BPT 的“备份”，使得可以较小增加处理器尺寸来存储附加分支预测信息。由于在一个处理器中既包含 L1 BPT 又包含 L2 BPT，则无须增加太多硬件成本便可同时提高分支预测的准确性和整体速度。另外，L2 BPT 容量的少量增加还可提高整体存储容量可以容纳具有大工作集的程序中的大量分支。

25 如果在 L1 BPT 中找到与初始 IP 地址有关的分支预测入口(称在 L1 BPT 中“命中(hit)”)，则使用与入口有关的分支预测信息来预测产生分支还是不产生分支(即预测后来的 IP 地址)。如果在 L1 BPT 中未发现与初始 IP 地址有关的分支预测入口(称在 L1 BPT 中“未命中

(miss)”), 则对与 IP 地址有关的指令进行解码来确定它是否为分支指令。如果是分支指令，则使用来自 L2 BPT 的分支预测信息来预测产生分支还是不产生分支。

对本发明的另一个实施例，软件给处理器提供提示以帮助分支预测。对这个实施例，用这些提示确定是否使用 L2 BPT 的分支预测信息或在软件自身中编码的分支预测信息来预测产生分支还是不产生分支。

可以发现这样的两层预测器设计对那些被指派执行大型应用程序任务的处理器特别有用，例如那些在服务器和工作站上运行的大型应用程序。下面更详细描述本发明的实施例，包括各种配置和实现。

图 1 是表示本发明方法的流程图。在步骤 205，搜索 L1 和 L2 分支预测表以查找与初始 IP 地址有关的分支预测入口。L1 BPT 是做了标记的关联表，而 L2 BPT 是未做标记的直接映射表。对一个实施例，L1 BPT 存储的分支预测入口少于 L2 BPT 存储的分支预测入口，但 L1 BPT 比 L2 BPT 快且具有比 L2 BPT 更高的分支预测准确性。因为 L1 BPT 存储较少的分支预测入口，因而 L1 BPT 只存储那些最近使用的、并因此最有可能在不久的将来再次使用的分支预测入口。

在图 1 步骤 210，确定 L1 BPT 中是否有与初始 IP 地址有关的命中。如果 L1 BPT 中有一个命中，则随后在步骤 215 确定 L1 BPT 中与初始 IP 地址有关的分支预测信息是否表明应该预测为产生分支。如果预测为产生分支，则在步骤 220 将后续 IP 地址重转向到 TAC 或 RSB 的地址。相反，如果预为不产生分支，则在步骤 225 不发送重转向信号。

如果在 L1 BPT 中没有命中(在 L1 BPT 中未命中)，则在图 1 的步骤 230 中，解码器对至少部分指令进行解码以确定与初始 IP 地址有关的指令是否为分支指令。如果确定指令为分支指令，则在步骤 230 中，解码器还对目标地址进行解码并判断指令是否为返回指令。对

于由分支指令把分支预测提示提供给分支预测器的实施例，在步骤 230 中解码器还对该提示信息进行解码。

如果在图 1 的步骤 235 中确定所述指令不是分支指令，则在步骤 240 不发送重转向信号。如果确定所述指令是分支指令，则下一步在步骤 250 确定与分支指令有关的命中是否为静态的。应该指出，对于本发明的不进行提示的另一个实施例，取消了步骤 250 和 260，且如果指令被确定为分支，则流程从步骤 235 跳到 255。

如果与分支指令有关的提示不是静态的(即是动态的)，则下一步在图 1 的步骤 255 中确定在 L2 BPT 中与初始 IP 地址有关的分支预测信息是否表明应该预测为产生分支。如果预测为产生分支，则在步骤 265 将后续 IP 地址重转向，预测后续 IP 为来自 TAC，RSB 的地址(如果指令被确定为返回指令)、或为解码器解码的地址。相反，如果预测不产生分支，则在步骤 240 不发送重转向信号。

相反，如果与分支指令有关的提示是静态的，则接下来在图 1 的步骤 260 确定提示是否表示应该预测为产生分支还是不产生分支。如果提示将产生分支，则流程进行到如上所述的步骤 265。相反，如果提示将不产生分支，则流程进行到如上所述的步骤 240。

应该指出，按照本发明的实施例，图 1 的步骤 220 或 225 中所做的分支预测比步骤 240 或 265 中所做的分支预测完成得早。

图 2 表示按照本发明实施例的分支预测流水线。按照图 2 的流水线，IP 复用器 10 的输出端提供初始 IP 地址给增量器 20、第一级(L1)分支预测器 21 和第二级(L2)分支预测器 22。增量器 20 适当地递增初始地址来产生后续的 IP 地址，并在第一流水线级期间把后续的 IP 地址送回到 IP 复用器 10 的输入端。IP 增量器 20 取到初始 IP 地址并按预定量进行递增。加到初始 IP 地址上的预定量是存储有正在执行的程序代码的连续指令或指令组的两个连续存储器地址之间的差值。

L1 分支预测器 21 可产生重转向信号并将该信号供给 IP 控制器

11 的输入端。这个信号表示该后续 IP 地址是否与初始 IP 地址顺序。如果 L1 分支预测器 21 向 IP 控制器 11 发送重转向信号，则这表明后续 IP 是非顺序的，且 L1 分支预测器 21 接着在第二级流水线级期间提供后续 IP 地址给 IP 复用器 10 的输入端。

5 L2 分支预测器 22 也可产生重转向信号并将这个信号供给 IP 控制器 11 的另一个输入端。这个信号同样表示后续 IP 地址是否与初始 IP 地址顺序。如果 L2 分支预测器 22 向 IP 控制器 11 发送重转向信号，这表示后续 IP 是非顺序的，且 L2 分支预测器 22 接着在第三级流水线级期间提供后续 IP 地址给 IP 复用器 10 的输入端。

10 然后，IP 控制器 11 基于来自 L2 分支预测器 22 和 L1 分支预测器 21 的信号来确定 IP 复用器 10 的三个输入中的哪一个可传到该复用器的输出端。如果 L1 分支预测器 21 和 L2 分支预测器 22 都没有发送重转向信号，则选择来自增量器 20 递增的 IP 地址作为复用器 10 输出的后续 IP 地址。如果 L1 分支预测器 21 发送重转向信号，则选择从 L1 分支预测器 21 输出的 IP 地址作为复用器 10 输出的后续 IP 地址。如果 L2 分支预测器 22 发送重转向信号，则选择从 L2 分支预测器 22 输出的 IP 地址作为复用器 10 输出的后续 IP 地址。

20 应该指出，这里使用的术语“初始 IP 地址”表示任意用做预测后续 IP 地址的参考点的 IP 地址。“初始 IP 地址”不意味着局限于与某具体软件应用程序的程序代码第一行有关的 IP 地址。初始地址可以是任何与应用程序的程序代码任意一行有关的 IP 地址。

25 图 3 表示按照本发明实施例的图 2 的 L1 分支预测器 21 和 L2 分支预测器 22。图 3 中的分支预测器可以完全包含在和处理器相同的半导体衬底上，分支预测器为该处理器进行分支预测。对于另一个实施例，分支预测器的一个或多个功能块位于单独的半导体衬底上。例如，为减少处理器的总体尺寸，L2 BPT 102 可置于单独的半导体衬底上。

 L1 BPT 100 是一个包括加有地址标记的分支预测入口的关联

表。每个地址标记都与包含分支指令的 IP 地址有关。在 L1 BPT 100 中的每个分支预测入口除了包括其有关的地址标记外还包括分支预测信息。分支预测器用这个分支预测信息来预测是否产生分支。存储在 L1 BPT 100 中的分支预测信息的具体类型相应于可由分支预测器执行的大量分支预测算法类型中的任何一种。其中大部分算法为本技术领域技术人员所了解。

例如，按照本发明的一个实施例，局部历史预测算法连同 L1 BPT 100 一起完成。对另一个实施例，在 L1 BPT 100 中完成全局历史分支预测算法或一个计数器预测器(即 2 位可逆计数器，也称为双(bimodal)分支预测器)。对另一实施例，L1 BPT 100 被分成两个或更多单独的分支预测表，每个表完成不同的分支预测程序。于是，选择器电路确定哪种算法为具体例子提供最准确的预测并选适合的表。

对于本发明的一个实施例，图 3 的 L1 BPT 100 是多路组相联。对于本发明的另一个实施例，L1 BPT 100 是完全相联。为提高 L1 BPT 100 中搜索分支预测入口的速度，保持表较小，使之大约有 512 到 2K 的分支预测入口的存储容量。

初始 IP 地址提供给图 3 的 L1 分支预测器 21 的输入端。这个初始 IP 地址被用来搜索 L1 BPT 100 及目标地址高速缓冲存储器(TAC)101。如果在 L1 BPT 中发现所述地址，则命中，且命中信号沿着命中/未命中信号线 121 发送到与门 144 的输入端。如果 L1 BPT 100 中未发现所述地址，则未命中，且未命中信号沿着命中/未命中信号线 121 发送出。如果 L1 BPT 100 中有命中，且在 L1 BPT 100 中有关的入口表示产生分支，则这由沿着产生/不产生信号线 122 发送到与门 144 的另一个输入端的产生信号来表示。如果 L1 BPT 100 中有命中，且在 L1 BPT 100 中有关的入口表示不产生分支，则这由沿着产生/不产生信号线 122 发送的不产生信号来表示。如果在 L1 BPT 100 中有产生的命中，且 L1 BPT 还表示分支是返回，则由沿着返回/不

返回信号线 143 发送到复用器 106 的控制输入端的返回信号表示。
如果在 L1 BPT 100 中有产生的命中，且 L1 BPT 表示分支不是返回，
则由沿着返回/不返回信号线 143 发送的不返回信号组来表示。

如果图 3 的 L1 BPT 100 中有命中，且 L1 BPT 表示产生分支，
5 则与门 144 输出重转向信号给 IP 控制器 11。如果在 L1 BPT 100 中
有未命中或在 L1 BPT 中有未产生命中，则与门 144 不输出重转向信
号。如果与门 144 输出重转向信号，则 IP 地址也从复用器 106 输出。
返回/不返回信号线 143 确定复用器 106 的输出。如果信号线 143 表
示分支是返回指令，则来自与复用器 106 输入端相连的返回堆栈缓
10 冲器的返回地址被传送到复用器 106 的输出端，且随后被传送到 IP
复用器 10。如果信号线 143 表示分支不是返回指令(对 L1 BPT 100
中产生的命中)，则来自通过目标总线 123 与复用器 106 的其它输入
端相连的 TAC 101 的目标地址被传送到复用器 106 的输出端，且随
后传送到 IP 复用器 10。另外，在 TAC 101 中发现的目标地址与命
15 中/未命中信号一起，也提供给 L2 分支预测器 22 的复用器 109 的一个
输入端，如下所述。

图 3 的 L2 BPT 102 是一个直接映射表，该表包括包含没有地址
标记的分支预测信息的分支预测入口。这个分支预测信息被分支预
测器用来预测是否将产生分支。存于 L2 BPT 102 中的分支预测信息
20 的具体类型相应于大量分支预测算法类型中的任意一种，这些分支
预测算法可由分支预测器执行，其中许多为本领域的技术人员所了
解。上面结合 L1 BPT 100 来描述这些分支预测算法中的某些分支预
测算法。L2 BPT 102 可以执行这些算法的任意一种或这些算法的任
何组合，而与 L1 BPT 100 所执行的算法类型无关。

25 特别是从成本角度看，图 3 的 L2 BPT 102 中执行的分支预测算
法占据小量空间是有利的。因此，根据本发明的一个实施例，L2 BPT
102 执行两比特计数器算法作为其分支预测方法。尽管两比特计数器
算法可能不像例如在 L1 BPT 100 中执行的局部或全局分支预测算法

那样准确，但使用两比特计数器算法的分支预测表使得每个分支预测入口仅需要两比特存储空间。执行或者局部或者全局分支预测算法的分支预测表使得每个分支预测入口可能需要多于两比特存储空间。通过使用未标记的分支预测入口和两比特计数器算法，L2 BPT 102 能存储大约 4 倍到 8 倍于或更多于 L1 BPT 100 在大约相同的空间中存储的分支预测入口数。这样，对本发明的一个实施例，L2 BPT 102 具有较大的约 2K 到 4K 或更多分支入口的存储容量。对于 L2 BPT 未被标记的一个实施例，没有有关静态提示的分支(下面更详细地描述)可以在退役(retirement)时更新 L2 BPT。

从图 3 的 L2 BPT 102 读取与初始 IP 地址有关的分支预测入口中存储的分支预测信息，且利用此信息来计算是否产生分支预测。根据分支预测，将产生或不产生信号沿 t/n 线 126 发送给复用器 107 的输入端。对 L2 BPT 102 是直接映射的本发明实施例，在这个表中总有命中。这意味着这些命中的某一百分比可能将一个 IP 地址与另外的 IP 地址的分支预测信息相联系。避免这个问题的一种方法是在 L2 BPT 102 中储存地址标记，并将这些标记与输入 IP 地址标记进行比较。然而，可以发现，通过不在 L2 BPT 102 中存储标记信息而减小表的大小的成本收益比通过存储标记来获得提高的分支预测信息准确性更有价值。

利用来自图 3 的 IP 复用器 10 的初始 IP 地址搜索输入指令缓冲器 103，并将有关的指令提供给指令解码器 104。对本发明的一个实施例，解码器 104 部分或完全地对指令进行解码来确定该指令是否为分支指令，并且，如果是，则解码器再确定目标地址并确定该指令是否为返回。根据指令是否为分支(或返回)，通过 b/nb 线 129 将合适的信号发送给与门 108 的输入端。把目标地址 130 从解码器 104 发送到复用器 109 的输入端。

对于指令中包括分支预测提示的实施例，图 3 的解码器 104 还确定分支预测提示是静态还是动态，并且，如果是静态，解码器确

定产生还是不产生提示。表示提示是静态还是动态的信号通过 s/d 提示线 127 提供给复用器 107 的控制输入端。通过 t/n 提示线 128 把表示提示是产生还是不产生的信号提供给复用器 107 的一个输入端。

静态提示向分支预测器表示：不管 L2 BPT 102 中发现的任何分支预测信息，有关的分支指令总是应该预测为产生或是不产生(根据产生/不产生提示值)。动态提示向分支预测器表示：L2 BPT 102 中发现的分支预测信息应用来把分支预测为产生或是不产生。对于本发明的另一个实施例，包含动态提示的指令还包括在初始执行分支指令时用作分支预测的产生/不产生提示。此后，储存在分支预测表中的分支预测信息被用来确定该分支指令的后续分支预测。

如果通过 s/d 提示线 127 提供给图 3 中复用器 107 的控制输入端的信号表示提示是动态的，则复用器选择通过 t/n 线 126 从 L2 BPT 102 来的产生/不产生信号作为其输出。如果提供给复用器 107 的控制输入端的信号表示提示是静态的，则复用器选择通过 t/n 提示线 128 从解码器 104 来的产生/不产生提示信号作为其输出。对于 L2 BPT 被标记的一个实施例，提示是动态的且在 L2 BPT 中有未命中，仍使用解码的预测。对这个实施例，来自 L2 BPT 102 的命中/未命中线(或其反向)可以和 s/d 提示线 127 (或其反向)相与(或与非)，输出连接到复用器 107 的控制输入端。

如果通过图 3 中的 b/nb 线 129 提供给与门 108 输入端的信号表示与 IP 地址有关的指令是分支(或返回)指令并且复用器 107 的输出值表示产生分支，则与门 108 输出重转向信号给与门 141 的输入端。否则，与门 108 不输出重转向信号。与门 141 的另一个输入端是连接到 L1 BPT 100 的命中/未命中信号线 121 的反向输入端。如果与门 108 的输出值表示重转向，则与门 141 便起到阻止重转向信号发送到 IP 控制器 11，但 L1 BPT 中有命中。这样做是因为可以发现来自 L1 BPT 100 的前期预测比来自 L2 BPT 102 的后期预测更准确。因此，当与门 108 的输出值表示再转向且 L1 BPT 100 中有未命中时，

与门 141 便输出重转向信号给 IP 控制器 11。

图 3 中的复用器 109 由通过返回/不返回(r/nr)信号线 140 来自指令解码器 104 的返回/不返回信号和通过命中/未命中(h/m)信号线 150 的 TAC 101 的命中/未命中信号控制。指令解码器 104 为复用器 109 选择合适的输出。例如, 如果 n/nr 信号线 140 表示指令是返回指令, 则来自返回堆栈缓冲器(RSB)142 的返回地址(提供给 L1 分支预测器 21 相同的 RSB 输出值)被传送给复用器 109 的输出端且随后被传送给 IP 复用器 10, 其中返回堆栈缓冲器 142 连接到复用器 109 的输入端。如果信号线 140 表示分支不是返回指令(产生分支的), 则如果 TAC 101 中有命中(如命中/未命中信号线 150 指示的)的话, 来自 TAC 101 的目标地址被传送给复用器 109 的输出端。否则, 由解码器 104 解码的且通过目标总线 130 提供给复用器 109 的另一个输入端的目标地址被传送给复用器 109 的输出端。

然后把后续的 IP 预测提供给处理器, 且由该处理器来执行位于那个地址的指令。稍后确定该预测正确或不正确。然后可刷新存储在 L1 BPT 100 和 L2 BPT 102 中的与预测的分支指令有关的分支预测信息以提高下一次执行分支指令的准确性。对于本发明的一个实施例, 如果与分支指令有关的 IP 地址在 L1 BPT 中未命中并且分支指令没提供静态提示, 则刷新 L1 BPT 以包括与 IP 地址有关的分支预测入口。

已参考本发明的具体示例性实施例描述了本发明。然而, 受益于本发明公开的人明白, 在不超出本发明更广的精神和范围的情况下可对这些实施例做出各种修改和改变。因此, 说明书与附图用于说明本发明而不是对本发明的限制。

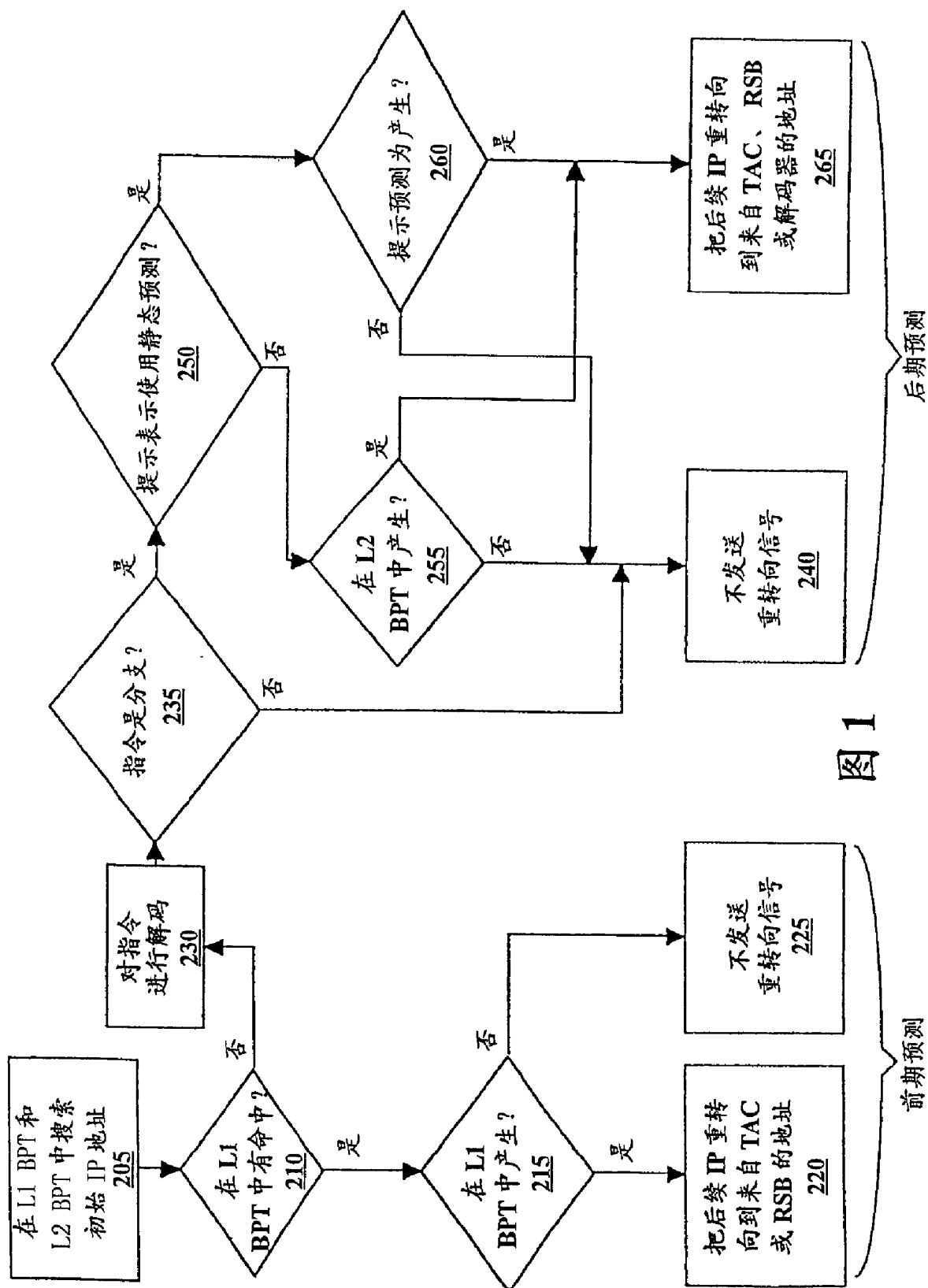
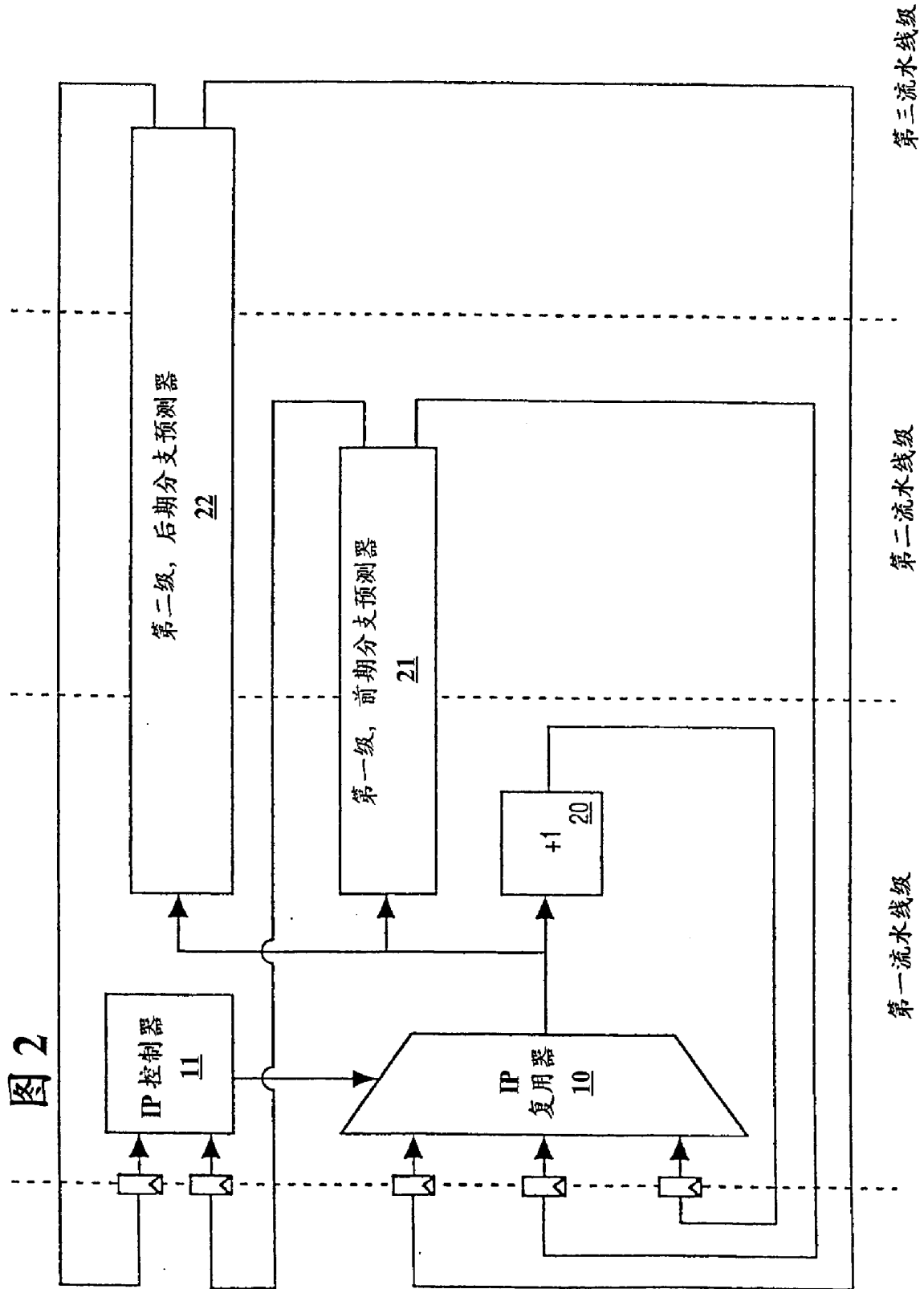


图 1



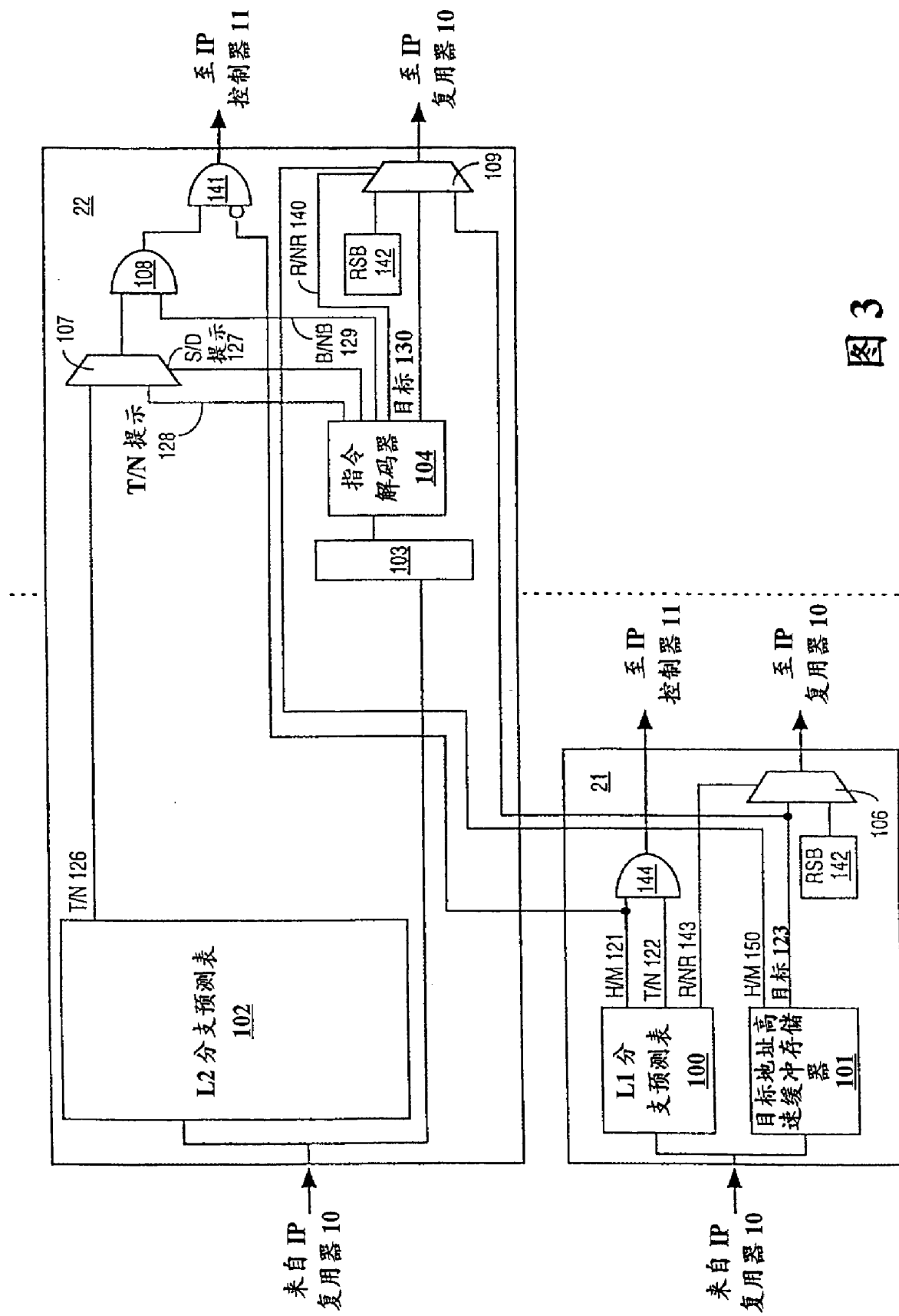


图 3